DOCUMENT RESUME

ED 217 058 TM 820 270

AUTHOR Daniel, Mark
TITLE Changes in English Vocabulary Scores, 1979-81, by Age and Region. Technical Report 1981-6.
INSTITUTION Johnson O'Connor Research Foundation, Boston, MA.
PUB DATE Sep 81
NOTE 13p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adolescents; *Age Differences; Comparative Analysis; *English; *Regional Characteristics; *Scores; Standardized Tests; *Verbal Tests; *Vocabulary Development; Young Adults

ABSTRACT

Age and regional differences in English Vocabulary scores are examined. The ages studied range from fourteen years through adulthood. The regions identified are Northeast, Southeast, Midwest, West, and Southwest. Scores were found to increase from age fourteen through the twenties, leveling off in the early thirties. The highest average increase was in the Northeast followed by Southeast, Midwest, West and Southwest. The statistical difference between the Northeast and Southwest was one third of a standard deviation. During the 1979-81 period, changes in vocabulary level have varied for different age groups and regions. The largest decline in four of the regions has been for examinees in their twenties. This population is weak in vocabulary skills when compared with other groups. If this interpretation is correct, a "dip" in curves should advance to older ages. Vocabulary levels have increased for teenage examinees in Northeast, Southeast and Midwest regions. This may be a result of increased attention to verbal skills in response to an observed decline in standardized test scores. The results suggest inter-regional differences in vocabulary level are increasing. The teenage vocabulary scores in the West and Southwest are declining. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED217058

CHANGES IN ENGLISH VOCABULARY SCORES, 1979-81, BY AGE AND REGION

by

Mark Daniel

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
A. Barker
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 270

JOHNSON O'CONNOR RESEARCH FOUNDATION
HUMAN ENGINEERING LABORATORY

Technical Report 1981- 6

September 1981

Changes in English Vocabulary Scores, 1979-81, by Age and Region

Mark Daniel

ABSTRACT: Age and regional differences in English Vocabulary scores are examined using a sample of 5,143 cases. Vocabulary scores increase from age 14 through the late twenties and level off in the early thirties. The Northeast averages highest, followed by the Southeast, Midwest, West, and Southwest; the difference between the highest and lowest regions is about a third of a standard deviation. During the period 1979-81 there have been changes in vocabulary level which have been different for different age groups and regions. In four of five regions, the largest decline (or smallest increase) has been for examinees in their twenties, and teenage examinees have shown an increase or slight decline. For teenage examinees vocabulary levels have increased in those regions that are already relatively high and have decreased elsewhere, suggesting that the inter-regional differences in vocabulary level are increasing rather than narrowing.

Introduction

As part of its standard battery of aptitude and knowledge tests the Johnson O'Connor Research Foundation administers a recognition test of English Vocabulary, Worksample 690 (Bowker, 1981). Age norms for this test were constructed in late 1979, using 2,627 cases administered during 1978 and 1979 (Daniel, 1979). Revised norms have now been constructed based on 5,143 cases administered between May 1979 and June 1981 (Brown, 1981). Although there is some overlap in these two norm samples, the amount of overlap is small, and a comparison of the two sets of data will give information on any changes in vocabulary level which may have occurred between 1978-79 and 1980-81.

The people tested by the Foundation have paid a fee to take the battery of tests. Their purpose in taking the tests is to gain information about their abilities which may help them choose a career or decide on a career change. The examinees tend to have more schooling than average; adults average about 16 years of education completed, and the parents of examinees have an average of 15.3 years of schooling (Bowker & Moore, 1981). Although the examinees are certainly a self-selected sample of the general population, there is no reason to believe that the nature of the sample has changed during the past three years; the testing fee and the number of examinees tested have remained approximately the same over that period. It should be noted that none of the students who have been tested under special grant programs are included in either of the data sets.

Relationship of Vocabulary Level to Age and Region

The relationship between age and vocabulary level is well known. Both the 1978-79 and 1979-81 data sets show a decelerating rate of vocabulary

growth during the teens and twenties, with a leveling-off starting in the early thirties, and with no apparent decline through about age fifty.

Table 1
Raw Data for Various Age Groups

1978-79				1980-81			
Ages	*N*	Mean	SD	Ages	*N*	Mean	SD
14	33	99.5	32.1	14	47	98.9	27.2
15	85	109.6	28.7	15	153	107.9	29.5
16	189	117.2	29.4	16	359	119.9	31.9
17	297	116.8	30.4	17	627	118.6	30.4
18	233	121.5	33.7	18	501	120.0	31.6
19	272	132.3	31.9	19	488	128.7	31.9
20	202	137.1	33.9	20	419	134.3	32.6
21	127	140.5	35.6	21	284	139.3	35.5
22	126	152.6	32.5	22	219	150.4	33.3
23	91	158.0	31.4	23	182	148.1	31.0
24	82	151.1	36.1	24	181	156.3	33.6
25-26	154	163.1	31.3	25-26	260	158.0	33.2
27-28	144	171.9	30.2	27-28	222	168.0	29.3
29-30	102	176.7	24.2	29-30	173	177.1	27.7
				31-32	165	173.9	28.9
31-33	94	176.2	28.4	33-34	146	182.4	27.0
34-36	93	178.0	27.5	35-37	175	179.0	30.8
37-40	90	179.6	27.8	38-40	151	180.8	30.0
41-45	92	183.2	23.7	41-45	174	177.6	28.0
46-50	66	187.0	26.3	46-50	130	185.1	27.4
51+	62	186.5	32.8	51+	87	186.9	29.7

Regional variation has been examined in the more recent data set, using multiple regression. Vocabulary scores are predicted by age, age-squared, and twelve dummy variables representing thirteen Foundation testing offices around the country. (A fourteenth office, in San Francisco, is new and has too few cases to permit analysis.) The unstandardized regression weights for the dummy variables tell how offices compare with one another in vocabulary level, in terms of raw scores on the vocabulary test, when the effects of age are controlled. These figures are presented in Table 2.

Table 2 indicates that, for a given age, the average New York examinee will score about ten and a half raw-score points higher than the average Boston examinee, and about twenty and a half points higher than the average Tulsa examinee. Given the overall standard deviation of raw scores of 39.73, the highest-scoring office is about a quarter of a standard deviation above the national average, and the lowest-scoring offices are about a quarter of a standard deviation below.

The inter-office variation is statistically highly significant. Using age and age-squared as predictors, 38.4% of the raw-score variance is ex-

plained; when the dummy variables representing the offices are added to the prediction equation, the predicted variance increases to 41.2% ($F(12,5108) = 20.3$, $p \ll .01$).

Table 2
Inter-Office Differences in Vocabulary Level

Office	*N*	Raw Score Difference[a]
New York	758	10.6
Washington, DC	321	5.7
Philadelphia	252	4.8
Chicago	476	3.8
Atlanta	355	1.9
Detroit	91	1.5
Boston	462	0.0
Los Angeles	422	- 2.9
New Orleans	62	- 3.3
San Diego	290	- 3.7
Houston	647	- 6.7
Fort Worth	787	- 9.0
Tulsa	200	- 9.9

[a]Unstandardized regression weights

Table 2 shows that vocabulary level is related to geographic region of the country. With the puzzling exception of Boston, the offices in the Northeast and Mid-Atlantic regions have the highest vocabulary levels. The Midwest offices (Chicago and Detroit) score somewhat lower, at about the same level as the Southeastern office (Atlanta). The two California offices are slightly below the median for all offices. The three offices in the Southwest are clearly the lowest in vocabulary level. Finally, New Orleans, which is geographically intermediate between the Southeastern office and the Southwestern offices, is also midway between them in vocabulary level.

Grouping Washington with Atlanta to form a numerically larger Southeast region, and deleting New Orleans because of uncertainty about which region it belongs to, we find the following inter-regional differences in vocabulary raw score (again, these are unstandardized regression weights with age and age-squared included in the prediction equation).

Table 3
Regional Differences in Vocabulary Level

Region	*N*	Raw score Difference
Northeast (Bost, NY, Phila)	1,468	7.3
Southeast (DC, Atlanta)	734	4.7
Midwest (Chi, Det)	565	4.6
West (LA, SD)	748	- 2.0
Southwest (Houst, FW, Tul)	1,628	- 7.4

Changes in Vocabulary Level, 1978-79 to 1980-81

Examination of Table 1 suggests that there has been a decline in vocabulary scores for examinees in their twenties, but that scores of examinees in their teens and over age 30 have remained about the same during the two-year period between the first and second norm construction. In this section of the report, these changes will be examined in more detail.

First it is necessary to evaluate the effect that changes in regional composition of the norms samples may have had on overall score levels. Table 4 shows the proportion of cases contributed by each office in each of the norm samples. Multiplying these proportions by the inter-office difference scores shown in Table 2 and summing across offices, gives an estimate of the effect of a change in regional composition. The conclusion is that the more recent sample is .24 of a raw-score point lower because of the change in the proportion of cases contributed by each office. Thus, the change in the composition has had very little effect on the level of scores.

Table 4
Sample Composition

	Proportion of Sample	
Office	1978-79	1980-81
Boston	.127	.089
New York	.166	.147
Philadelphia	.033	.049
Washington, DC	.019	.062
Atlanta	.096	.068
Chicago	.079	.092
Detroit	.023	.018
Los Angeles	.145	.082
San Diego	.042	.056
Houston	.127	.125
Fort Worth	.097	.153
Tulsa	.046	.039
San Francisco		.007
New Orleans		.012

In order to discover how changes in vocabulary level between 1978-79 and 1980-81 may vary with age and region, a series of multiple-regression analyses have been performed in which the dependent variable is vocabulary-score change and the predictors are age, age-squared, and four dummy variables representing the five regions in Table 3. The dependent (change) variable is the difference, for each examinee in the 1980-81 sample, between his raw score and the mean raw score for his region and age group in the 1978-79 sample. These 1978-79 mean scores are shown in Table 5, along with the 1980-81 means for comparison. Only ages 16 and older are used in the analyses. Also, ages 21-25 are transformed to age 23, ages 26-30 to 28, and ages 31+ to 33.

Table 5
Means by Age and Region

		1978-79		1980-81		
	Age	*N*	Mean	*N*	Mean	SD
Northeast	16	80	122.6	113	123.4	31.1
	17	99	115.0	185	120.2	31.8
	18	64	123.6	120	130.8	31.0
	19	85	137.2	128	136.3	33.7
	20	52	142.2	106	140.1	33.7
	21-25	179	159.6	289	157.3	34.3
	26-30	114	177.4	161	176.0	30.1
	31+	158	180.6	317	187.1	28.5
Southeast	16	14	112.1	52	123.5	33.3
	17	26	122.8	64	125.4	34.8
	18	18	139.8	60	129.3	36.3
	19	47	140.5	51	136.5	36.4
	20	30	143.4	70	142.8	30.7
	21-25	67	147.7	141	149.4	35.8
	26-30	38	178.9	74	169.6	27.4
	31+	48	181.2	141	181.5	32.3
Midwest	16	21	115.3	58	129.4	34.2
	17	40	118.2	79	118.1	28.8
	18	30	125.9	48	120.2	31.0
	19	25	125.7	50	129.1	31.8
	20	16	138.6	36	131.1	28.6
	21-25	31	147.6	103	157.5	29.1
	26-30	37	172.3	75	171.4	28.5
	31+	55	179.9	86	184.5	25.6
West	16	26	121.7	33	114.9	31.8
	17	48	119.5	71	120.2	27.9
	18	47	112.2	68	111.9	26.5
	19	35	126.9	59	127.1	25.7
	20	33	141.7	46	131.8	34.4
	21-25	96	153.9	129	145.9	31.6
	26-30	65	164.9	89	164.6	31.1
	31+	115	179.7	183	179.0	26.1
Southwest	16	48	108.3	94	108.9	28.2
	17	84	114.8	215	114.3	28.5
	18	74	119.4	197	112.0	29.1
	19	80	126.8	191	122.2	30.2
	20	71	128.3	151	128.5	31.4
	21-25	121	137.7	321	140.0	31.2
	26-30	71	166.9	116	159.7	31.6
	31+	126	176.5	277	170.5	28.3

Note: Standard deviations are not available for the 1978-79 sample.

The results of the first analysis are shown in Table 6 below. There is no linear relationship between age and vocabulary change. However, the combined effect of age and age-squared is significant, meaning that there is a curvilinear relationship between age and vocabulary change. The effect of region also is statistically significant, signifying that there is a greater change in vocabulary scores in some regions than in others.

Table 6
Regression Analysis

Variables	R^2	F	df	p
Age	.0000			
Age, Age^2	.0014	3.41	2,4841	<.05
Region	.0055	6.70	4,4841	<.01
Age, Age^2, Region	.0069			

The complete regression equation using unstandardized weights is:

$$Diff' = 22.58 - 2.15(Age) + .0429(Age^2) + 4.53(NE) + 1.92(SE) + 6.06(MW) + 0.13(SW)$$

(Note that the Western region, not coded, has an implicit regression weight of zero.)

It is possible that the relationship between age and vocabulary change is different in different regions. To investigate this, regression analyses were performed using interaction terms: age x region, and age^2 x region. This was done in two steps. The simplest result would be to find different linear relationships between age and vocabulary change in different regions. This could occur, even though the first regression analysis showed no overall linear relationship between age and vocabulary change, if the relationships within different regions had different slopes which canceled one another out. Therefore, a first analysis looked solely at the ability of age, region, and age x region to predict vocabulary change.

Table 7
Regression Analysis with Interactions Involving the Linear Component of Age

Variables	R^2	F	df	p
Age, Region	.0055			
Age, Region, A x R	.0066	1.34	4,4838	NS

Because the linear interaction terms do not explain a significant amount of the variance in vocabulary change scores, a second analysis was performed using both the linear and quadratic interaction terms.

Table 8
Regression Analysis with All Interactions

Variables	R^2	*F*	*df*	*p*
Age, Age^2, Region	.0069			
Age, Age^2, Region, A x R, A^2 x R	.0111	2.57	8,4838	<.01

There is significant variation among the regions in the shape of the relationship between age and vocabulary change. These relationships are shown in Figure 1, based on the complete regression equation:

$$Diff' = 33.71 - 3.24(Age) + .0672(Age^2) + 30.29(NE) + 14.27(SE) - 17.90(MW) - 60.49(SW) - 2.28(A \times NE) - 0.76(A \times SE) + 2.13(A \times MW) + 5.53(A \times SW) + .0468(A^2 \times NE) + .0096(A^2 \times SE) - .0443(A^2 \times MW) - .1172(A^2 \times SW)$$

For each region, this equation reduces to:

$$NE: Diff' = 64.00 - 5.52(Age) + .1140(Age^2)$$
$$SE: Diff' = 47.98 - 4.00(Age) + .0768(Age^2)$$
$$MW: Diff' = 15.81 - 1.11(Age) + .0229(Age^2)$$
$$SW: Diff' = -26.78 + 2.29(Age) - .0500(Age^2)$$
$$W \quad Diff' = 33.71 - 3.24(Age) + .0672(Age^2)$$

Discussion

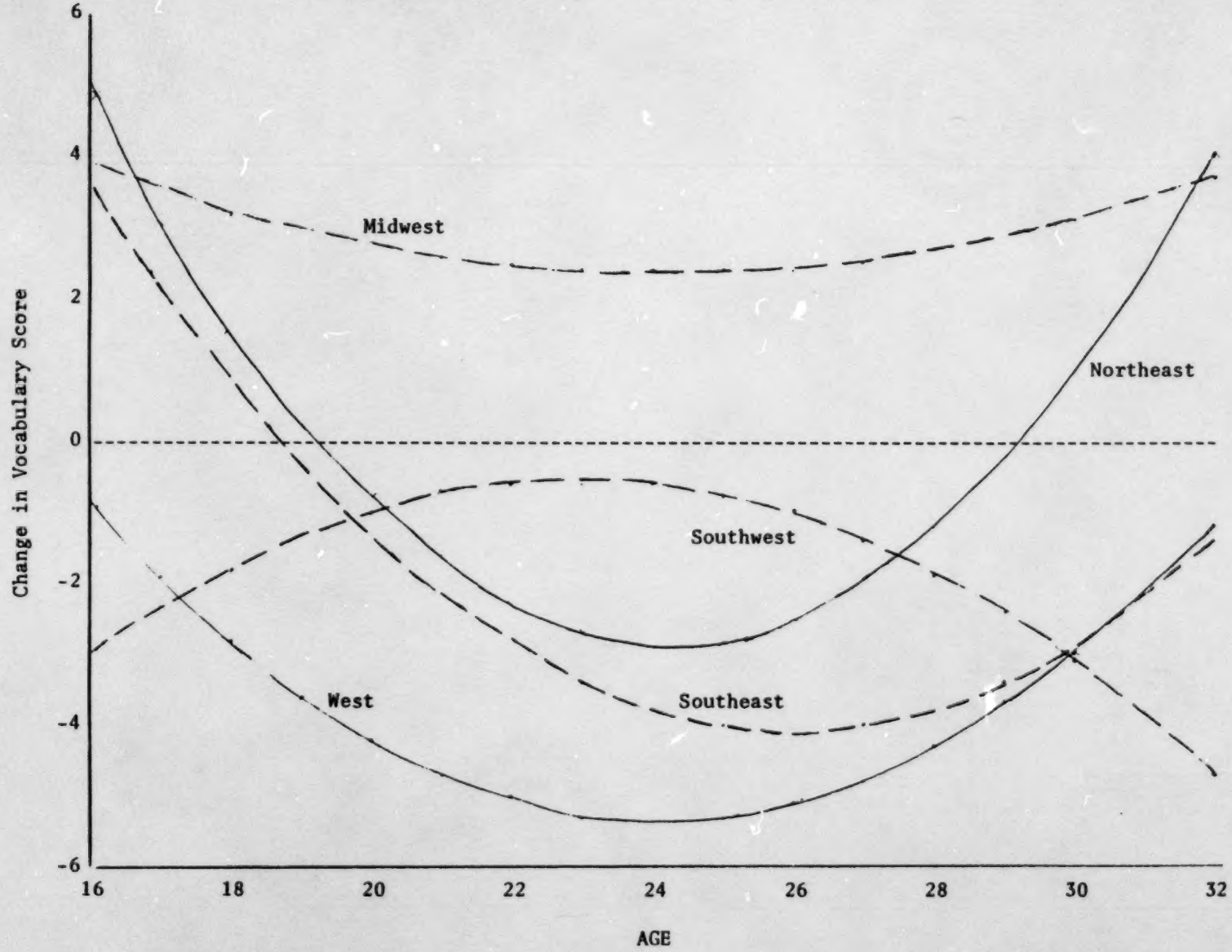
For three of the regions--the Northeast, Southeast, and West--examinees in their mid- to late-twenties show the sharpest decline, from three to five and a half raw-score points. Teenagers in the Northeast and Southeast show an increase in vocabulary level; in the West, their vocabulary levels decline, but not as greatly as for examinees in their twenties. A fourth region, the Midwest, displays a nearly-constant increase in vocabulary scores for all ages, but even here examinees in their twenties show less of an increase than teenage examinees.

The "dip" in four of the five regional age curves may indicate that examinees now in their mid-twenties are particularly weak in vocabulary compared with people several years older or younger. This age group would have graduated from high school between the late 1960s and the mid-1970s. If this interpretation is correct, as time goes on we should see this "dip" in the curves advance to older ages.

These age curves indicate that teenagers are improving in vocabulary in the Northeast, Southeast and Midwest. Perhaps this reflects the increased attention that is being paid to verbal skills in response to the observed decline in SAT and other standardized test scores.

The Southwest shows an age curve which is the opposite of those in the other regions. Here, teenage examinees show a greater decline than examinees in their twenties. The implications of this curve are the opposite of those for the other regions: the decline in vocabulary appears to be getting worse,

Figure 1
Change in Vocabulary, 1979-81, by Age and Region



in the sense that the youngest examinees are showing the largest decline.

The trends shown in Figure 1 suggest that the gap between the high-vocabulary and low-vocabulary regions of the country is getting larger rather than smaller. In the Northeast, Southeast and Midwest the vocabulary levels of teenagers appear to be increasing, whereas in the West and Southwest teenage vocabulary scores are declining.

REFERENCES

Bowker, Richard. English vocabulary manual. Boston: Johnson O'Connor Research Foundation, 1981.

Bowker, Richard & Moore, Gene. Demographic information concerning Foundation examinees (Statistical Bulletin 1981-37). Boston: Johnson O'Connor Research Foundation, 1981.

Brown, Diane. Revised norms for English Vocabulary Worksample 690 A&B (Statistical Bulletin 1981-39). Boston: Johnson O'Connor Research Foundation, 1981.

Daniel, Mark. Construction of English Vocabulary norms. Unpublished manuscript, Johnson O'Connor Research Foundation, 1979.

APPENDIX
Means and Standard Deviations for
Cases Used in Regression Analysis of Change
(N = 4848)

Variable	Mean	SD
Vocab. Change	-0.38	31.1?
Age	23.18	6.03
Age^2	573.61	301.67
Northeast	.29	.46
Southeast	.13	.34
Midwest	.11	.31
Southwest	.32	.47
Age x NE	6.86	11.18
Age x SE	3.16	8.32
Age x MW	2.50	7.36
Age x SW	7.25	11.01
Age^2 x NE	172.07	315.09
Age^2 x SE	79.17	229.20
Age^2 x MW	60.36	196.38
Age^2 x SW	173.73	300.35